ANSI/NISO Z39.xx-199x

Status:

A Draft American National Standard
Developed by the
National Information Standards Organization

Version 1.1
February 16, 1999

About NISO Draft Standards

This is a draft standard and subject to change. To learn the current status of this proposed standard contact the NISO Office: 4733 Bethesda Avenue, Suite 300, Bethesda, MD 20814 USA; Telephone: 301-654-2512; Fax: 301-654-1721; Internet nisohq@niso.org.

NISO standards are developed by the Standards Committees of the National Information Standards Organization. A rigorous review process includes offering each NISO voting member and other interested parties an opportunity to review the proposed standard. In addition, approval requires verification by the American National Standards Institute that its requirements for due process, consensus, and other criteria for approval have been met by NISO; NISO standards thus also become American National Standards.

This standard may be revised or withdrawn at any time. Purchasers may receive current information on all NISO standards by calling or writing the National Information Standards Organization.

Address inquiries to NISO, 4733 Bethesda Avenue, Suite 300, Bethesda, MD 20814; nisohq@niso.org; Telephone: 301-654-2512; Fax: 301-654-1721.

To Order Any NISO Standard Contact:

NISO Press Fulfillment, P.O. Box 338, Oxon Hill, MD 20750-0338 USA
Telephone: 1-800-282-6476 or 301-567-9522

Version 1.1
February 16, 1999

Contents

Foreword
Background
Goals of Syntax Committee
Elements of the DOI and Metadata
The DOI in Context (DOI System, Handle™ System, IDF)
Character Set Issues

1. Introduction
1.1 Purpose
1.2 Scope

2. Standards and References
2.1 Referenced Standards

3. Definitions

4. Format and Characteristics of the DOI
4.1 Prefix
4.2 Suffix

5. Maintenance Agency

APPENDIX A

APPENDIX B

Designation of Maintenance Agency

APPENDIX C

Examples

APPENDIX D

Related Standards and References
Secondary Standards
References to Standards in Development:
References:

APPENDIX E

Application Issues
Encoding Requirements
Encoding Characters
Mandatory/Recommended Encoding
Table 1 – Mandatory Encoding
Table 2 – Recommended Encoding

Foreword

This Foreword is not part of the American National Standard for the Digital Object Identifier (DOI), ANSI/NISO Z39.xx-199x. It is included for information only.

This standard defines the syntax for the Digital Object Identifier (DOI). The Digital Object Identifier (DOI) system is an identification system for intellectual property in the digital environment. Developed by the International DOI Foundation on behalf of the publishing industry, its goals are to provide a framework for managing intellectual content including activities such as linking users with content owners, facilitating electronic commerce, and enabling automated copyright management.

Background

Some history on the development of this draft standard is needed to understand how and why the draft standard appears in its present form.

DOI System Background

The Internet is a new environment for information transactions, and requires new enabling technologies to provide services and to protect intellectual property. Systems must be developed to identify, authenticate, and protect content to insure that what the user is requesting is what is being delivered. At the same time, the rights owner of the information must be sure that copyright in content is respected and protected.

In considering the new systems required, international publishers realized that a first step would be the development of an identification system to be used for intellectual property in the digital environment. Such a system was launched at the Frankfurt Book Fair in October 1997: the Digital Object Identifier (DOI) system provides a unique identification mechanism for content in all media, and also a way to link users of the materials to the rights holders or their agents to facilitate automated digital commerce.

NISO Involvement and Role of the International DOI Foundation (IDF)

The ongoing management of the DOI System is handled by an international, not-for-profit, membership-based organization called the International DOI Foundation with offices in the United States and Geneva, Switzerland. The Foundation is responsible for licensing Directory Managers/Registration Agencies and technology providers, for setting policy for the system, and for encouraging development of the related enabling technologies to build the infrastructure for electronic transaction systems such as copyright management.

Though the DOI System was originally developed by the publishing industry, it was recognized that the DOI system would have a broader scope and that it should work with established standard bodies as much as possible. ANSI/NISO involvement was requested in early 1998 in order to recognize the broader potential use of a digital object identifier, and it was decided to develop a standard for the syntax of the DOI identifier string.

Goals of Syntax Committee

The standards subcommittee was established with the following goals:

*Formalize the syntax for the DOI identifier string to enable DOI registration.
It has been possible to register DOIs since 1998 but there has been concern that the syntax has not been conclusively set.

*Determine the elements of the DOI identifier string.
Several proposals to add elements to the DOI identifier string had been proposed and the subcommittee was tasked to determine which ones should be included in the string.

*Limit the scope of the standard to the DOI identifier string.
The DOI system is made up of a number of parts including the identifier string, the resolution mechanism, and the Directory. This standard addresses only the syntax for the identifier string.

While limiting itself to the DOI identifier string, the Syntax Committee took into account the wider context of the DOI system, the Handle™ system, the International DOI Foundation and the Internet. This Foreword and the Appendices provide important information and references for understanding the DOI system, how the DOI is being used, how it relates to other standards and the Internet, and where to get more information.

Elements of the DOI and Metadata

From the earliest development of the Digital Object Identifier (DOI), one of the most widely discussed issues has been whether or not the identifier string of which the DOI is composed should be meaningful. The syntax of this string, as defined by this standard, contains a set of components of very limited meaning. They are as follows:

Each DOI string begins with a code indicating that within the Handle System of the Corporation for National Research Initiatives (CNRI) the string will be resolved by the Local Handle System reserved for the DOI.

The next component of the DOI string notes the prefix assigned to the DOI Prefix holder that originally created the DOI. It should be noted that this prefix provides no information about current ownership of the object that the DOI string references.

The final part of the DOI string contains an unspecified identifier.

Thus, any particular DOI string contains no information about the object identified, and is in that sense meaningless. Because of this decision not to provide any information about the object referenced within the syntax of the DOI, the string will be of little use in isolation without accompanying information about the object identified. Such information about objects is called metadata and may be aggregated in metadata databases.

The following recommendations are accordingly included here:

1. No DOI string should be registered without an accompanying set of metadata describing the object being referenced in the syntax string.
2. The DOI Syntax Standard Maintenance Agency should provide the latest information about the relevant metadata schemas and any databases that aggregate metadata about DOI referenced objects.

The DOI in Context (DOI System, Handle™ System, IDF)

The DOI System is an implementation of the Handle™ System, developed by the Corporation for National Research Initiatives (CNRI). The DOI System is managed by the International DOI Foundation, which sets policies, appoints service providers and ensures the successful operation of the System. The IDF has issued a document outlining the DOI issues entitled "Guidelines for the Issuance and Use of DOI" (the most current version will be available at http://dx.doi.org/10.1000/25).

Character Set Issues

The Committee discussed character sets and decided that Unicode and UTF-8 will be used in this standard. The SICI standard limited characters to 7-bit ASCII but the committee felt that since the XML specification (Extensible Markup Language 1.0 - http://www.w3.org/TR/REC-xml) and the underlying Handle™ technology (Internet Draft -- Handle System: A Persistent Global Naming Service - Overview and Syntax - http://hdl.handle.net/4263537/4006) support Unicode and UTF-8 (The Unicode Consortium - http://www.unicode.org/) that the DOI Syntax should follow suit.

The only reserved character in the DOI string is percent (%). Percent is used for hex encoding DOIs and therefore must always be hex encoded as %25. There are issues (addressed in Appendix E) concerning character encoding when using DOIs in URLs. Along with percent (%), space (SP), quotes (") and hash (#) must be encoded when they are used in URLs or HTML pages. However, there is no requirement for them to be encoded when registering DOIs and they are stored in their unencoded form in the DOI Directory.

Users of ASCII should note that ASCII is UTF-8 compliant. Thus, DOIs consisting entirely of characters taken from the ASCII character set require no modifications or changes (other than those detailed in Appendix E) in order to meet this standard's requirement for Unicode and UTF-8 compliance.

ASCII characters in the DOI string are case insensitive.

NISO Voting Members
[list]

NISO Board of Directors
[list]

Standards Committee xx

NISO acknowledges with thanks and appreciation the contributions of the following persons who served on NISO Standards Committee xx that drafted the proposed standards for the Digital Object Identifier syntax.

[list]

Digital Object Identifier

1. Introduction

1.1 Purpose

ANSI/NISO Z39.xx-199x defines the syntax for a unique and persistent character string called the *Digital Object Identifier (DOI)*.

1.2 Scope

This standard defines the syntax of the DOI character string, which is made up of a prefix and suffix and is registered according to the requirements set by the International DOI Foundation (IDF). Policies governing the assignment and use of DOIs are determined by the IDF and are outside the scope of this document.

2. Standards and References

The Referenced Standards are those that are required to construct a DOI. Secondary Standards and References includes citations to documents that can be of use in conjunction with the DOI.

2.1 Referenced Standards

Unicode Consortium "The Unicode Standard Version 2.0" ISBN: 0-201-48345-9 (http://www.unicode.org/)

See Appendix C for related standards and references.

3. Definitions

Digital Object Identifier (DOI)
A unique character string used in a system conforming to the rules of, and deposited in the directory administered by, the IDF.

Directory
A repository in which DOIs are deposited and attendant locations are maintained.

Directory manager
The entity which controls the Directory.

DOI prefix
The Directory and the Registrant codes issued by the Registration agency to a Registrant for use as the prefix in the DOIs allocated by that Registrant.

DOI suffix
The character string assigned by a Registrant. The suffix shall be unique within the range of DOIs specified by the DOI prefix held by the Registrant.

International DOI Foundation (IDF)
The body set up to support the needs of the intellectual property community in the digital environment by establishing and governing the DOI System, setting policies for the System, appointing service providers for the System, and overseeing the successful operation of the System.

Registration Agency [DOI registration agency]
One of the agencies appointed by the International DOI Foundation to register and allocate DOI prefixes to Registrants, and which subsequently accepts DOIs being deposited by Registrants. A Registration Agency may also be a Directory Manager.

Registrant
An organization or entity that has requested and been allocated one or more DOI prefixes by a Registration Agency.

Deposit
The act of entering into the Directory a DOI and associated information necessary for the DOI to be used.

Registration
The act of allocating the DOI prefix to a Registrant by the Registration Agency.

4. Format and Characteristics of the DOI

The DOI is made up of the *Prefix* and the *Suffix*. Within the prefix are the Directory Code <DIR> and the Registrant Code <REG>. The suffix is made up of the DOI Suffix String <DSS>. See Appendix A for the allowed length of the DOI string and a statement on case sensitivity.

The syntax of the DOI string is:

<DIR>**.** <REG> /<DSS>

Detailed explanations of each section of the DOI string follow.

4.1 Prefix

<DIR> - Directory Code (required) –

See Appendix A for all valid values for the Directory Code. The Maintenance Agency is

responsible for updating the list of valid values.

<REG> Registrant's Code (required)
Separated from <DIR> by ".". This is the number assigned to the registrant by the Registration Agency.

DOI Prefix Character Set

<DIR><REG> = any octets that map to UTF-8 encoded Unicode 2.0 characters except octets %x2E and %x2F which map to ASCII characters '.' '/'. Octet %x25, which maps to '%', must always be encoded as %25.

4.2 Suffix

<DSS> DOI Suffix String (required)
The unique string assigned by the Registrant.

DOI Suffix Character Set

<DSS> = any octets that map to UTF-8 encoded Unicode 2.0 characters. Octet %x25, which maps to '%', must always be encoded as %25.

5. Maintenance Agency

The Maintenance Agency designated in Appendix B shall review suggestions for new data elements, interpret the rules prescribed by this standard, and maintain a listing of inquiries and responses that may be used for potential future enhancement of this standard. Questions concerning the implementation of this standard and requests for information should be sent to the maintenance agency.

APPENDIX A

This appendix is not part of the American National Standard for Digital Object Identifier (DOI), ANSI/NISO Z39.xx-199x. It is included for information only.

Valid values for Directory Code (See Section 4.1), Maximum Length and Case Sensitivity

Valid value for <DIR> = 10
Maximum length of entire DOI string = 128 characters
ASCII characters in the DOI string are case insensitive.

The above information is kept up-to-date by the Maintenance Agency (See Appendix B).

APPENDIX B

Designation of Maintenance Agency

This appendix is not part of the American National Standard for Digital Object Identifier (DOI), ANSI/NISO Z39.xx-199x. It is included for information only.

Maintenance Agency Administered by the International DOI Foundation

The functions assigned to the Maintenance Agency as specified in Section 5 will be administered by the International DOI Foundation (http://www.doi.org/). Questions concerning the implementation of this standard and requests for information should be sent to:

L'Association Internationale DOI
Attn: Norman Paskin
3, avenue de Miremont
NW CH-1206 Geneva
Switzerland
TEL: 0041.22.830.10.80
FAX: 0041.22.830.10.81
E-mail: doi@worldcom.ch

The International DOI Foundation
Attn: Norman Paskin
1718 Connecticut Avenue, 7th Floor
Washington, DC 20009, USA
TEL: 202-232-3335 ext. 228
FAX: 202-745-0694
E-Mail: n.paskin@doi.org

APPENDIX C

Examples

This appendix is not part of the Digital Object Identifier (DOI), ANSI/NISO Z39.xx-199x. It is included for information only.

A DOI for the Authors' Licensing and Collecting Society's Byline service is:

10.054/1418EC1N2LE

A DOI (incorporating a SICI) from an article in the *Journal of the American Society for Information Science*, published by John Wiley & Sons, is:

10.1002/(SICI)1097-4571(199806)49:8<693::AID-ASI4>3.0.CO;2-O

An example of a DOI from the *European Physical Journal C*, published by Springer Verlag, is:

10.1007/s100529901036

The DOI for the article "ABO Blood Group System" from *Encyclopedia of Immunology 2e Online*, published by Academic Press, is:

10.1006/rwei.1999.0001

APPENDIX D

Related Standards and References

This appendix is not part of the Digital Object Identifier (DOI), ANSI/NISO Z39.xx-199x. It is included for information only.

Standards for item identification have been proliferating in recent years. The referenced standard cited in Section 2 is required for the construction of the DOI syntax. This appendix includes Secondary Standards and References to Standards in Development, which cite documents that provide information that may be useful when using DOIs. Other references provide additional information on the DOI.

When American National Standards cited below are superseded by a revision, the revision shall apply.

Secondary Standards

ANSI X3.4:1986 *American National Standard for Information Systems* – Coded

Character Sets – 7-bit American National Standard Code for Information Interchange (7-bit ASCII)

ANSI/NISO Z39.9-1992, International Standard Serial Numbering (ISSN) ISO 3297:1986

ANSI/NISO Z39.23-1997, Standard Technical Report Number
http://www.niso.org/stantech.html#z3923

ANSI/NISO Z39.56-1996, Serial Item and Contribution Identifier (SICI), Version 2
http://sunsite.berkeley.edu/SICI

ISO 2108:1992, Information and Documentation -- International Standard Book Numbering (ISBN)

ISO 3901:1986, Information and Documentation -- International Standard Recording Code (ISRC)

ISO 10957:1993, Information and Documentation --International Standard Music Number (ISMN)

ISO/IEC 10646, Universal character code
http://www.unicode.org/unicode/standard/standard.html

Uniform Resource Identifier
http://info.internet.isi.edu:80/in-notes/rfc/files/rfc2396.tx

References to Standards in Development:

NISO SCAP, Book Item and Contribution Identifier (Draft 5, January 1999)

ISO/TC 46/SC9 N 252, Committee Draft ISO/CD 15706, Information and Documentation -- International Standard Audiovisual Number (ISAN)

ISO/TC 46/SC9 N 253, Committee Draft ISO/CD 15707, Information and Documentation -- International Standard Work Code (SWC) -- Part 1: Musical Works

References:

Guidelines for the Issuance and Use of DOI Version 3.1
http://dx.doi.org/10.1000/25

Document Object Model
http://www.w3.org/DOM/

Dublin Core Metadata Initiative
http://purl.oclc.org/dc/

Handle System
http://www.handle.net/

Internet Draft -- Handle System: A Persistent Global Naming Service - Overview and Syntax (submitted to the IETF 14 Nov 97; Updated 16 Jul 98)
http://hdl.handle.net/4263537/4006

Publisher Item Identifier
http://www.elservier.nl/homepage/about/pii

APPENDIX E

Application Issues

This appendix is not part of the Digital Object Identifier (DOI), ANSI/NISO Z39.xx-199x. It is included for information only.

DOIs will be placed within URLs and transmitted via HTTP to be resolved. The URL currently used for the resolution of DOIs is http://dx.doi.org/

If a DOI were used as a link on an HTML page, the code could look like this:

<A HREF=http://dx.doi.org/10.1006/rwei.1999.0001**>10.1006/rwei.1999.0001</A>**

Encoding Requirements

There can be special encoding requirements when a DOI is used with URLs and transmitted via HTTP. The syntax for Uniform Resource Identifiers (URIs) is much more restrictive than the syntax for the DOI. A URI can be a Uniform Resource Locator (URL) or a Uniform Resource Name (URN). There will be instances when certain characters contained in DOIs will need to be encoded to be transmitted within a URL.

Encoding Characters

Since URL syntax conventions are more restrictive than DOI conventions, hexadecimal (%) encoding must be used for characters in a DOI that are not allowed, or have other meanings, in URLs. Hex encoding, or escaping characters, consists of substituting for the given character its hexadecimal value preceded by percent. Thus, # becomes %23 and

http://dx.doi.org/10.1000/456#789 is encoded as http://dx.doi.org/10.1000/456%23789. The browser does not now encounter the bare #, which it would normally treat as the end of the URL and the start of a fragment, and so sends the entire string off to the DOI network of servers for resolution, instead of stopping at the #. **NOTE:** The DOI itself does not change with encoding, merely its representation in a URL. A DOI that has been encoded is decoded before being sent to the DOI Registry. At the moment the decoding is handled by the proxy server http://dx.doi.org/. Only unencoded DOIs are stored in the DOI Registry database. For example, the number above is in the DOI Registry as "10.1000/456#789" **and not** "10.1000/456%23789".

DOI creators and maintainers need to keep the following rules in mind:

There are few character restrictions for DOI number strings per se. When DOIs are embedded in URLs, they must follow the URL syntax conventions, but the same DOIs need not follow those conventions in other contexts, e.g. inventory databases, DOI Registry.

Always present DOIs in URLs in their encoded form to end users.

The percent character (%) must always be hex encoded (%25) in any web form, batch input, or URL. Other excluded characters must always be hex encoded in URLs, but do not need to be encoded when registering the DOIs through the web forms or batch mode.

Mandatory/Recommended Encoding

Table 1 shows a list of characters that are not allowed within the URL syntax: these native characters must therefore be replaced by hex-encoding. Table 2 shows additional characters where it is recommended that characters be replaced by hex-encoding. The distinction between Tables 1 and 2 is between practical experience with current web browsers and the more formal specification of URL syntax. The web is still a relatively new environment and both the formal specifications and web browser implementations are something of a moving target. It can say with certainty that the characters in Table 1 must be hex encoded. It cannot say with certainty that the characters in Table 2 must or must not be hex encoded, only that the formal specifications recommend it but no recent set of browsers require it.

The tables are formatted to show the given character, the encoding required by the DOI system for input, the encoding required for use of a DOI within a URL, and finally what is stored. The | symbol is used to show a series of acceptable alternatives, e.g., the fact that # can be entered into a web form as either # or %23 is shown as # | %23. Note that all the given characters must be hex encoded within URLs but that only %, because of its use as marking the start of a hex encoding, must be hex encoded in the various DOI forms.

Table 1 – Mandatory Encoding

Character	Input		URLs	DOI Database
	Web Form	Batch		
%	%25	%25	%25	%
"	" | %22	" | %22	" | %22	"
#	# | %23	# | %23	%23	#
SP	SP | %20	SP | %20	%20	SP

Table 2 – Recommended Encoding

Character	Input		URLs	DOI Database
	Web Form	Batch		
<	<| %60	< | %60	%60	<
>	> | %62	> | %62	%62	>
{	{ | %123	{ | %123	%123	{
}	} | %125	} | %125	%125	}
^	^ | %94	^ | %94	%94	^
[	[| %91	[| %91	%91	[
]	] | %93	] | %93	%93	]
`	` | %96	` | %96	%96	`
|	"|" | %124	"|"| %124	%124	|
\	\ | %92	\ | %92	%92	\